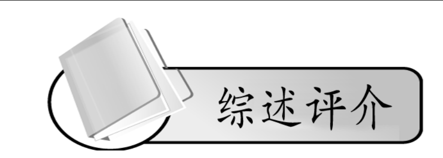


数据监管的政策研究综述*

张梦霞 [1,2]　**顾立平** [1]
[1](中国科学院文献情报中心　北京 100190)
[2](中国科学院大学　北京 100049)

摘要: 【目的】探索数据监管政策的实施细节, 促进数据监管政策的建立。【方法】通过梳理国内外数据监管的相关理论成果, 并分别从科学数据的遴选标准、存储规范和传播交流机制三方面进行归纳和总结。【结果】数据监管的政策要素主要包括: 数据遴选标准(符合数据提交流程要求、优先遴选原则、数据真实可用的声明、数据来源不具争议性); 数据存储规范(遵循相关政策、保障数据完整、满足通用技术标准、保证长期可持续发展); 传播交流机制(遵循法律法规和行业指南、开放获取传播许可协议、传播行为的免责声明、数据重用的说明文件)等。【局限】需要结合我国的实际, 进一步完善政策框架中的细节。【结论】科研组织、协会、资助机构等应该积极推进和制定数据监管政策。

关键词: 数据监管　数据管理服务　数字档案管理　数据权益管理　数据遴选　长期保存
分类号: G302

1　数字监管下的科学数据具有高质量重用价值

以数据论文的形式共享科学数据为数据发现、重用以及科研人员的成果获得领域内学者的认可和信赖奠定了基石[1]。在 e-Science 环境中, 人们对数据驱动科研所需要的、具有高质量保障的数据集有更高要求[2]。为此, 科学界普遍认为需要在数据生命周期内进行主动积极的监管, 以便数据更符合新环境下对其利用和重用的要求, 同时更好地促进学术社群内数据的共享与交流[3]。数字监管指在科学数据的生命周期内采取主动的方式(数据标注、评价、选择、转换等)对数据进行有效管理, 使其增值, 从而提供更广泛的数据共享[4]。

美国学院与研究图书馆协会(ACRL)在 2012 年将数据监管作为图书馆界的主流发展趋势之一[5], 国内外图书情报界开展一系列有关数据监管的举措[6-8], 包括:

(1) 构建数据知识库, 国际上目前建设较好的数据知识库有 Edinburgh DataShare、Dryad、figshare 等。

(2) 开展数据监管教育。美国伊利诺伊大学香槟分校(UIUC) 2006 年开设了数据监管课程。此后, 北卡罗来纳州立大学、密歇根大学等也先后开设数据监管课程[9]。

(3) 开发基于生命周期的数据管理计划服务工具, 如美国加利福尼亚数字图书馆的数据管理计划在线工具(DMP Online)。

(4) 建立数据管理基础设施, 开展数据监管实践, 如约翰霍普金斯大学的数据保护(Johns Hopkins University's Data Conservancy)、普渡大学科研数据库(Purdue University Research Repository)、罗格斯大学的 RUresearch 数据门户(RUresearch Data Portal)和美国康奈尔大学图书馆的 DataStar 项目。

在推进一系列工作取得成绩的同时, 也出现若干新的挑战[10], 如用以监管的数据质量控制问题、科研机构数据政策的制定、用以支持数据管理者进行数据监管的参考指南等, 随着数据监管工作的不断开展,

通讯作者: 顾立平, ORCID: 0000-0002-2284-3856, E-mail: gulp@mail.las.ac.cn。
*本文系中国科学院项目"中国科学院文献出版领域择优支持人才"(项目编号:院1434)的研究成果之一。

这些问题将会被不断放大。

数据监管已成为图书馆、档案馆的一项重要发展战略，其实施过程中会涉及一系列政策问题，例如，用以监管的数据需要符合哪些标准要求？如何对数据进行存储？数据传播交流的机制是什么？等等。本研究针对上述问题进行政策调研工作，探讨数据监管的政策要素，以期在实践工作中为数据监管的政策制定与决策服务提供参考。

2 数字监管下科学数据管理规范的研究框架设计

在数据监管框架下，需要对科学数据进行鉴别、遴选、存储、评估、分析、再利用、共享，这不仅仅是工程问题，更涉及政策问题，将上述服务所涉及的权益关系进行研究提问(Research Inquiry)，应该包括：

(1) 科研人员进行科学活动时所产生的科学数据，所需要的管理程序是什么？

(2) 经过管理程序初步遴选的科学数据，在存储的数据基础设施上，应有哪些管理方式？

(3) 学术社群对科研人员提交并存储在数据知识库中的科学数据进行传播共享时，应该具有怎样的共识公约？

据此，本文的研究框架分别从“研究问题”和“观测问题”两方面进行论述，如表 1 所示：

表 1　研究框架

研究问题	观测问题
数字监管下的科学数据需要符合哪些遴选标准?	(1)数据是否符合提交流程要求？数据集是否涵盖应当保存的数据内容？元数据是否符合规范？ (2)哪些数据需要进行优先遴选？数据内容是否是一次性生成的原始数据？是否具有保存价值的数据？ (3)数据是否真实可用？数据是否可验证和可重用？是否存在虚假信息？ (4)数据来源应该符合什么要求？数据是否存在法律或伦理限制？
数字监管下的科学数据如何进行存储?	(1)数据存储是否有可遵循的政策？政策应该具有哪些规定？ (2)数据是否被完整地存储？有无防止篡改或无意改动的措施？数据是否进行安全备份？ (3)数据集是否符合通用的技术标准？对数据格式、软硬件具有怎样的要求？ (4)数据保存是否具有可持续性？是否具有可持续的资金投入？如何应对不可抗拒的事故？
科学数据传播交流应该遵循怎样的机制?	(1)数据是否遵循法律指南或行业行为准则？ (2)如何解决数据传播中的知识产权问题？是否获得开放获取传播的授权许可协议？ (3)数据涉及人类主体问题时，如何进行规避？是否获得传播行为的免责声明？ (4)数据是否有支持共享和重用的说明文档？是否有能说明数据结构、数据访问和引用规范的说明文件？

3 科学数据的遴选标准

国际科学和技术信息理事会(International Council for Scientific and Technical Information, ICSTI)信息生命周期流程框架[11]中将遴选作为整个工作流程不可或缺的一个环节。数据监管同样离不开数据的遴选，其原因包括：

(1) 数据备份和镜像站维护需要一定成本，所以需要遴选具有保存价值的数据。

(2) 如果数据不进行遴选，所存储的数据可能会不断增长、泛滥甚至造成冗余，从而让数据的发现、挖掘、利用变得困难[12]。

(3) 若干具有保存价值的数据随着科研项目生命周期的结束可能会造成流失，良好的遴选实践标准能及时保存这类数据。

(4) 科研资助条款和科研教育机构对数据保存期限有所要求，数据遴选提供一种保存期满后是否继续存储的鉴定评估途径。

3.1 数据应符合提交流程要求

数据监管至少有三种实施情况[13]:

(1) 数据贡献者产生、描述、自行提交存储数据。

(2) 数据监管人员采集、评估、遴选、存储和长期保存数据。

(3) 数据贡献者提交后，数据监管人员进行筛选、审核、管理与提供利用。

普遍做法是采用第三种方式，双方按照一套工作流程(Workflow)实施。

数据贡献者和数据监管人员需要共同遵守的原则:

(1) 数据集应该涵盖需要保存的内容，保存内容的完整丰富性对数据理解、利用和监管具有重要影响。

(2) 元数据应该尽可能详尽完整，才能供其他人高效率使用。

3.2　数据优先遴选原则

在同等条件下，数据监管不仅要确保数据能长期保存，更重要是实现数据知识的交流，优先遴选可信、可用和具有价值的数据更能符合这一目标[14]。优先遴选原则包括:

(1) 一次性原始数据。数据的记录是短暂的或一次性的，如天气观测、火山爆发或降雨记录等。

(2) 不可重复产生的数据。观测对象原来存在，但观测的变量可能会随着时间的推移而出现不同的情况，例如短时间内受测者(人类/动物)依然可以重复测验，但当时实验所获得的数据是不可重复的。

(3) 非冗余数据。指计算机系统内不存在重复或无用的数据。

(4) 优先考虑具有科学研究、历史记录和社会经济价值的数据。

数据监管人员在实施上述原则的第 4 项时，需要有可操作的概念界定。有学者综述美国科学研究理事会(National Research Council, NRC)、DDC、ANDS 等的报告[10,15]指出:

(1) 数据的科学价值指数据支持科学活动，对科技成果具有可验证性。

(2) 历史价值指数据不仅提供给科研人员所用，更包括社会团体和个人使用。

(3) 社会价值指数据能反映当代社会的利益，并在未来的重用中，有助于社会经济的发展。

3.3　数据真实可用的声明

数据贡献者应当提交一份表明所提交数据真实可用的声明。其含义包括:

(1) 数据可解读。

(2) 数据可被验证和重用，即数据能对研究成果进行有效验证，并通过数据追踪能确认科研结论的可信度。

(3) 数据集不存在伪造信息，可以作为科研成果之一，并且作为证明相关结论的依据。

(4) 无刻意筛选屏蔽信息。

数据贡献者提交数据真实可用的声明一方面表示数据贡献者愿意对此负责，另一方面可以帮助数据监管人员快速理解数据。

3.4　数据来源不具争议性

数据监管的第一原则是合法性，包括数据监管的依据、过程和内容，所以，用以监管的数据不应当以不合法、不合理、不道德的方式产生。数据监管应该避免存在法律、道德、伦理的争议，例如数据所有权、涉及人类主体的实验、涉及个人信息、国家安全、保密以及数据提供者要求的共享前提[16]。上述原则要求澄清以下几点:

(1) 数据生成过程中不违反学术道德、科学伦理和现有法律法规。

(2) 学术活动产生的科学数据应该遵守法律、伦理、社会实践等的共同规范，即便是特定学科的信息管理准则也将此类原则视为优先考虑。

(3) 数据传播应考虑各方利益关系人的合法权益。

行业规范对这些原则的具体实施带来重要参考价值[17]：以晶体学领域而言，数据使用者据此开发新产品、开展新研究和申请新项目时，要与数据贡献者先行沟通，取得正式或非正式的同意/参与许可；在地理生物学(Geobiology，也称为地球生物学[18])领域，国家资助的观测仪器和科研项目所产生的科学数据普遍要求开放共享，大型研究所和受委托的公司则会尽量保留自己的权利，在开放获取时滞期上提出不同要求。

4　数据存储规范

数据基础设施包括但不限于[19]:

(1) 大型仪器设备及其信息平台;

(2) 专业领域的数据交流网络;

(3) 数据中心：大数据资源基地或项目仓储;

(4) 数据银行(Data Bank)：基于协议或条款为数据贡献者提供数据存储的知识库;

(5) 数据档案馆：为感兴趣的终端用户群体提供特定数据资源的数据知识库;

(6) 图书馆：电子文件平台、数据知识库和机构知识库。

尽管这类基础设施建设目的不同，但在面临数据监管挑战时，利益相关人(科研资助者、科研管理者、项目负责人、数据贡献者、数据监管人员、数据使用者)需要达成一致共识或遵守一套公约。

4.1　优先遵守资助机构和科研机构的数据政策

科研资助者的资助政策、条款及管理办法中对项目负责人如何处理数据做了相关规定。作为支持数据

监管的基础设施，自然也要遵守、尊重、遵循他们的数据政策，才能要求、邀请、邀约数据贡献者提交符合数据遴选标准的数据，并被数据监管人员审核、审查、审定能否提供给数据使用者，包括科研团体和公众。

根据前期调研成果[20-21]，科研资助机构和科研教育机构要求受到资助的科研项目所产生的数据在遵守机构管理办法下进行长期保存和开放共享。这类政策针对：

(1) 学科范围；

(2) 提交时间，一般要求资助项目完成后6-12个月内进行提交存储；

(3) 最低保存期限，科研数据至少保存3年以供查验，安全保存10年以供利用；

(4) 开放获取日期，科研成果正式发布后的12个月内实施开放获取；

(5) 存储地点，在机构知识库或者第三方可以证明具有保存与传播功能的，并能保护各方权益的数据中心等地方存储。

4.2 保障数据的完整性

数据完整性是衡量数据质量的重要标准之一，不同于“数据真实可用的声明”(见3.3节)中数据贡献者有关数据真实完整的声明，对数据基础设施所要求的数据存储规约而言，强调的是严格保障数据完整性的具体措施原则，例如，美国政治与社会研究校际联盟(Inter-university Consortium for Political and Social Research, ICPSR)基于OAIS模型开发了社会科学数据知识库，该知识库为保证数据完整性，特别强调考虑到数据背景信息、保存描述信息，以及在数据访问上的用户权限问题，这些措施在一定程度上保证了数据在知识库中的完整性。

这类完整性的基本要求包括：

(1) 防止篡改

假如数据需要修改，应有三种选项：

①为了让数据使用者能明确不同版本的数据，数据贡献者进行数据修改、更新、调整时，应有一个说明和记录。

②数据监管人员因为格式统一、复制保存等的需要或其他原因在小范围内修正数据时，需要告知数据贡献者或事前应有明确规范。

③当接受到第三方提出数据集在存储、传播以及内容上有问题的异议时，应由数据贡献者、数据监管人甚至科研管理人员进行论证。

上述三种情况都应该保证数据不能被任何人以任何理由不经过当事人同意、许可的情况下篡改。

(2) 防止无意改动

以数据监管而言，唯有数据错误时，无论该数据来自于提交时已经有的瑕疵、数据格式迁移时造成的数据损失、或其他意外所造成的数据破坏等，才会再次校订、补充和更正数据。

(3) 安全备份

为长期保存、防止自然、意外灾害或人员组织变动等情况所造成的数据流失，根据合理使用原则，对数据进行复制或迁移措施。

4.3 满足通用的技术标准

数据监管有赖于一个适合的技术标准，其理由包括：

(1) 数据迁移时，双方如果采用统一标准，会提高工作效率，降低经济成本[22]；

(2) 数据备份时，所用的工具、方法、操作程序及专业人员技能等相似，可以减低数据噪音；

(3) 数据转换时，应避免仅由少数厂商或者技能小组实行，增加实施方案的可选性。

通用技术标准的功能一般包括：文件格式、类似OAIS的参考标准、永久保存的唯一标识符以及支持远程访问、存储和验证的标准。其特性主要包括：机器可读、人可识别、容易访问、可转换格式以及兼容上述条件的开放性，主要考虑到数据监管需要对数据进行定期迁移转换，具有开放性的技术标准是有弹力、可适应多种技术策略和经济有效的[23]。

4.4 保证长期可持续发展

数据监管的核心目的之一是维持高质量数据的开放共享，避免重要数据的流失。数据开放获取能够让人们充分利用数据、进行溯源和承认科研贡献，因此需要长期可持续的发展机制保证这类记录真实存在。

数据监管的长期可持续发展有三项针对性：

(1) 非一次性项目工程。数据管理通常会面临有价值的科学数据容易随着项目结束而流失，因此，一个长期可持续的数据监管机制是较好的解决方案；

(2) 不以盈利为目的的商业经营模式。数据监管具有收入支出效益，且高标准要求所遴选的数据具有客观、完整和一致性等不受利益因素干扰的前提条件，长期可持续发展方案会优先考虑非商业经营模式；

(3) 遇不可抗事故的应变措施。当受到自然灾害或人为事故等不可抗因素时，需要有一套标准作业程序应对这类事故带来的损失[24]。

5　传播交流机制

数据监管的重要目标之一是为用户提供科学数据的开放获取[25]。数据基础设施从任务流程(数据提交、组织、处理、存储和共享等)考虑并建立数据监管前期阶段的内部管理机制，但在真正推进数据共享时，会面临着各方面阻碍，如没有明确告知数据贡献者数据利用情况，如何去尊重、保护数据贡献者权益等。从政策规划角度而言，完整的数据监管除了遴选规范、存储规约之外，传播交流机制也是整体政策方案中的重要组成部分[26-27]。

5.1　遵循国家法律法规或行业的最佳实践标准

在考虑各种传播交流机制的优先次序上，首先是国家的法律法规，其次是行业的最佳实践标准。具体而言，可以包括:

(1) 遵循著作权及其邻接权的合法权益[28](数据贡献者理当享有的精神权利，以及财产权利的分配)、知识产权转移转化(IPCs)[29](科研资助者和科研管理者对于产生市场价值的数据提供免费获取的要求)、单位和/或知识库政策等;

(2) 科学记录管理规范，如《澳大利亚负责任的研究行为准则》[30]、《英国研究理事会良好科研行为管理的行为准则和政策》[31]等(对科学数据存储、存储时间、如何存储及提供共享等做了规定)。

5.2　开放获取授权许可协议

许可协议(License)是著作权拥有者对如何使用内容进行的规范性说明，开放获取授权许可协议是数字化、网络化环境中如何充分保障数据贡献者权益以及赋予数据使用者权利的重要依据。目前，比较通用的协议是 CC-BY 要求数据使用者尊重(数据贡献者)作者精神权利，对数据贡献者进行署名(说清楚用谁的数据)。此外，还有CC0等协议，可以对政府统计数据等客观事实不需要署名具体数据贡献者的情况进行授权许可。

由于数据的采集、处理、筛选和符合遴选标准的提交过程涉及一定程度的智力劳动，因此，数据集普遍以 CC-BY 或者更加严格的许可协议进行使用授权。而元数据是描述这类事实(数据集)的数据，建议以 CC0 的方式授权许可。由此，数字监管规划可以明确传播使用许可，以 CC-BY 授权数据集、以 CC0 授权元数据的方式实施。

5.3　传播行为的免责声明

数据传播共享中主要伦理问题就是涉及个人或组织的敏感数据以及数据的安全控制问题[32]，而传播行为的免责声明是告知数据使用者经数据监管人员遴选、存储和提供开放获取的数据具有一定品质，虽然数据贡献者会对数据来源不具争议性进行说明，但数据监管人员在审核数据后，有义务和责任告知数据使用者关于数据的正当来源和使用方的正当使用途径。例如，涉及人类主体数据需要知情同意签名文件；涉及敏感数据或政治数据需要伦理委员会批准证明；涉及第三方合同的数据需要授权同意书。

传播行为的免责声明并不仅仅是尽到告知义务，而且对促进良好学术交流氛围、促进自我约束的道德习性、增进开放获取良性循环等具有参考意义和实践价值。

5.4　数据重用的说明文件

综上所述，无论是遴选标准、存储规约还是传播交流等 11 项政策主旨，其最终目标都是为了让科学数据具有良好的管理和传播使用。若要达成上述各项政策，数据监管的最后一关是数据重用的说明文件，解释若要用这批数据进行重复结果、重置成品、重新创造时，如何理解它们的结构、字段名含义等。这类说明文件主要有：数据字典；与数据创建环境相关的背景信息，如项目性质、数据采集与处理方式等；数据访问与引用的推荐方式等。

6　结　语

6.1　数据监管中的数据管理规范的政策要素总结

针对本文所提出的三个研究问题归纳为可操作的观测问题及其政策要素，包括：数据遴选标准、数据存储规约和传播交流机制三大管理重点，如表 2 所示。

6.2　实践意义

在 e-Science 环境中，研究流程和出版流程的讨论较多，监管流程固然有若干讨论数字监管的重要成果[33-35]，但多是从技术、系统和教育培训等角度讨论，而从政策角度，特别是针对数据的政策，国内还在初步探索中。

表 2　数据监管中科学数据管理规范的政策要素

管理重点	政策要素
数据遴选标准	(1)符合数据提交流程要求(数据集涵括需保存的内容、元数据符合规范并且详尽) (2)优先遴选原则(一次性原始数据、不可重复产生的数据、非冗余数据、优先考虑具有科学研究、历史记录和社会经济价值的数据) (3)数据真实可用的声明(可解读、可验证、可重用、无虚假伪造信息、无刻意筛选屏蔽信息) (4)数据来源不具争议性(数据生成过程及学术活动产生的数据应不违反学术道德、科学伦理和现有的法律法规；数据传播考虑到各方利益者的合法权益)
数据存储规范	(1)优先遵守资助机构和科研机构的数据政策(学科范围、提交时间、最低保存期限、开放获取日期、存储地点等) (2)追求数据的完整性(防止篡改、防止无意改动，以及安全备份等) (3)满足通用技术标准(机器可读、人可识别、容易访问、可转换格式，兼容上述条件的开放性) (4)保证长期可持续发展(非一次性项目工程、不以盈利为目的的商业经营模式、遇不可抗事故的应变措施)
传播交流机制	(1)遵循国家法律法规以及行业最佳实践指南(著作权及其邻接权、知识产权转移转化、单位和/或知识库政策以及科学记录管理规范) (2)开放获取传播授权许可协议(无特殊情况下，元数据以 CC0 方式，数据集以 CC-BY 方式授权许可) (3)传播行为的免责声明(涉及人类主体数据需要知情同意签名文件；涉及敏感数据或政治数据需要伦理委员会会批准证明；涉及第三方合同的数据需要授权同意书) (4)数据重用的说明文件(数据字典、项目性质、数据采集与处理方式、数据访问与引用推荐方式)

数据监管是一项系统工程，涉及数据对象及其完整性、数字监管的技术措施、法律和组织因素以及其他如政策标准、开放规范、元数据等要素[36]。因此，本文在前人研究基础上，做适当的补充，并且所梳理的研究框架(表1)以及政策框架(表2)涵括了各个利益相关人的基本权益，可为政策制定提供参考依据。

6.3　研究局限

表 1 研究框架下所列的问题，还需要结合实际情况进行具体细化，重点分析当前数据监管存在的现实问题。表 2 所梳理的可操作执行的数据监管方案，在实施过程中还要配合资助机构、科研教育机构和信息服务机构的数据管理政策。在此基础上，为有效推进这方面的工作，还需要结合科学数据管理计划(DMP)共同实施，可以取得良好成效。

将本研究结果进行推广时，需要注意结合实际案例研究，在制定有关管理规范时，需要顾及到科研资助者、科研管理人员、科研项目负责人、科研人员等的实际需求。

6.4　未来研究

对科研团队进行实地调研，对数据监管的研究中，除了数据的遴选方面，还可以对数据管理计划进行政策要素分析，对数据与衍生数据的权益管理进行研究，开放获取传播授权许可协议，与科研管理部门进行沟通协调，结合本研究制定一套完整的流程方案。

(致谢：感谢中国科学院文献情报中心王辉老师、欧阳峥峥老师对论文从学科馆员的角度给予指导，并对论文中存在的问题和不足给予解答和建议。感谢中国科学院文献情报中心组织编译《如何评价及挑选用于管理的研究数据》，为本文的研究提供参与。编译人员：慕慧鸽、王璐，以及吴蓉对所收集参考文献的规范化整理。)

参考文献：

[1] 刘晶晶，顾立平. 数据期刊的政策调研与分析——以 Scientific Data 为例[J]. 中国科技期刊研究，2015，26(4): 331-339. (Liu Jingjing, Ku Liping. The Policy Research and Analysis of Data Journals: Taking Scientific Data as an Example [J]. Chinese Journal of Scientific and Technical Periodicals, 2015, 26(4): 331-339.)

[2] Rombouts J, Princic A. Building a 'Data Repository' for Heterogenous Technical Research Communities Through Collaborations [EB/OL]. [2015-06-14]. http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1014&context=iatul2010.

[3] Shreeves S L, Cragin M H. Introduction: Institutional Repositories: Current State and Future [J]. Library Trends, 2008, 57(2): 89-97.

[4] 张智雄，吴振新，刘建华，等. Digital Curation 和 Digital Preservation 之概念辨析[J]. 现代图书情报技术，2014(1): 4-13. (Zhang Zhixiong, Wu Zhenxin, Liu Jianhua. Analysis of the Difference Between Digital Curation and Digital Preservation[J]. New Technology of Library and Information Service, 2014(1): 4-13.)

[5] McLure M, Level A V, Cranston C L, et al. Data Curation: A

Study of Researcher Practices and Needs [J]. Portal: Libraries and the Academy, 2014, 14(2): 139-164.

[6] Wright S J, Kozlowski W A, Dietrich D, et al. Using Data Curation Profiles to Design the Datastar Dataset Registry [J]. D-Lib Magazine, 2013, 19(7-8). http://www.dlib.org/dlib/july13/wright/07wright.html.

[7] 宋秀芬, 邓仲华. 数据监护的知识技能与教育研究[J]. 图书馆学研究, 2014(21): 5-11. (Song Xiufen, Deng Zhonghua. The Knowledge and Skills Needed for Data Curation and Education Research [J]. Research on Library Science, 2014(21): 5-11.)

[8] 曹霞. 国内 Data Curation 研究现状与热点分析[J]. 图书情报工作, 2014, 58(18): 144-148. (Cao Xia. Research on Development of Studying on Data Curation in China [J]. Library and Information Service, 2014, 58(18): 144-148.)

[9] 高红文, 陈清文. 国外数据监管研究综述及启示[J]. 图书馆学研究, 2013(10): 2-4, 27. (Gao Hongwen, Chen Qingwen. Foreign Data Curation Research Review [J]. Research on Library Science, 2013(10): 2-4, 27.)

[10] 王芳, 慎金花. 国外数据管护(Data Curation)研究与实践进展[J]. 中国图书馆学报, 2014, 40(4): 116-128. (Wang Fang, Shen Jinhua. Advances in Data Curation Abroad: Research and Practice [J]. Journal of Library Science in China, 2014, 40(4): 116-128.)

[11] Hodge G M. Best Practices for Digital Archiving: An Information Life Cycle Approach [J]. D-Lib Magazine, 2000, 6(1). http://www.dlib.org/dlib/january00/01hodge.html.

[12] Whyte A, Wilson A. How to Appraise & Select Research Data for Curation [EB/OL]. [2015-06-14]. http://www.dcc.ac.uk/sites/default/files/documents/How%20to%20Appraise%20and%20Select%20Research%20Data.pdf.

[13] Higgins S. Draft DCC Curation Lifecycle Model [J]. International Journal of Digital Curation, 2008, 2(2): 82-87.

[14] Constantopoulos P, Dallas C, Androutsopoulos I, et al. DCC&U: An Extended Digital Curation Lifecycle Model [J]. International Journal of Digital Curation, 2009, 4(1): 34-45.

[15] Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources [M]. Washington: National Academies Press, 1995: 33-42.

[16] Carlson J, Fosmire M, Miller C C, et al. Determining Data Information Literacy Needs: A Study of Students and Research Faculty [J]. Portal: Libraries and the Academy, 2011, 11(2): 629-657.

[17] Cragin M H, Palmer C L, Carlson J R, et al. Data Sharing, Small Science and Institutional Repositories [J]. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 2010, 368(1926): 4023-4038.

[18] "地球生物学前沿"专题[EB/OL]. [2015-05-12]. http://blog.sciencenet.cn/blog-528739-807103.html. (The Latest "Earth Biology Frontier" Project [EB/OL]. [2015-05-12]. http://blog.sciencenet.cn/blog-528739-807103.html.)

[19] Beagrie N, Greenstein D. A Strategic Policy Framework for Creating and Preserving Digital Collections: A Report to the Digital Archiving Working Group [R]. London: British Library Research and Innovation Centre, 1998:17-18.

[20] 张瑶, 顾立平, 杨云秀, 等. 国外科研资助机构数据政策的调研与分析——以英美研究理事会为例[J].图书情报工作, 2015,59(6): 53-59. (Zhang Yao, Ku Liping, Yang Yunxiu, et al. Research Data Policies of Research Funding Agencies: Case Study of British and American Research Councils [J]. Library and Information Service, 2015,59(6): 53-59.)

[21] 杨云秀, 顾立平, 张瑶, 等. 国外科研教育机构数据政策的调研与分析——以英国 10 所高校为例[J]. 图书情报工作, 2015, 59(5): 53-59. (Yang Yunxiu, Ku Liping, Zhang Yao, et al. Research on Data Policies of Research and Education Institutions: Case Study of British Universities [J]. Library and Information Service, 2015, 59(5): 53-59.)

[22] Hedstrom M. Research Issues in Migration and Long-term Preservation [J]. Archives and Museum Informatics, 1997, 11(3): 287-292.

[23] Mellor P, Wheatley P, Sergeant D. Migration on Request, a Practical Technique for Preservation [A]. // Research and Advanced Technology for Digital Libraries[M]. UK: Springer Berlin Heidelberg, 2002: 516-526.

[24] Barateiro J, Antunes G, Freitas F, et al. Designing Digital Preservation Solutions: A Risk Management-Based Approach [J]. International Journal of Digital Curation, 2010, 5(1): 4-17.

[25] Hedstrom M, Lampe C. Emulation vs. Migration: Do Users Care? [J]. RLG DigiNews, 2001, 5(6): 5-11.

[26] Knight G, Hedges M. Modelling OAIS Compliance for Disaggregated Preservation Services [J]. International Journal of Digital Curation, 2008, 2(1): 62-72.

[27] McMeekin S M. With a Little Help from OAIS: Starting down the Digital Curation Path [J]. Journal of the Society of Archivists, 2011, 32(2): 241-253.

[28] Besek J M. Copyright Issues Relevant to the Creation of a Digital Archive [J]. Microform & Imaging Review, 2003,

32(3): 86-97.
[29] Bearman D. Intellectual Property Conservancies [J]. D-Lib Magazine, 2000, 6(12). http://www.dlib.org/dlib/december00/bearman/12bearman.html.
[30] Australian Code for the Responsible Conduct of Research [EB/OL]. [2015-06-08]. http://www.nhmrc.gov.au/publications/synopses/r39syn.htm.
[31] RCUK Policy and Code of Conduct on the Governance of Good Research Conduct [EB/OL]. [2015-06-12]. http://www.rcuk.ac.uk/review/grc/default.htm.
[32] Kirschenbaum M G, Ovenden R, Redwine G, et al. Digital Forensics and Born-Digital Content in Cultural Heritage Collections [M]. Washington: Council on Library and Information Resources, 2010:49-50.
[33] 谢婧, 陈雅. 国内外科学数据监管研究[J]. 大学图书情报学刊, 2014, 32(4): 114-119. (Xie Jing, Chen Ya. Research on the Curation of Scientific Dada Home and Abroad [J]. Journal of Academic Library and Information Science, 2014, 32(4): 114-119.)
[34] 沈婷婷, 卢志国. 科研项目不同阶段的科学数据监管方法[J]. 图书馆建设, 2013(3): 49-51. (Shen Tingting, Lu Zhiguo. Science Data Curation Methods at Different Stages of Scientific Research Projects [J]. Library Development, 2013(3): 49-51.)
[35] 蔚海燕, 卫军朝. 国外高校数据监管项目的调研与分析[J]. 图书情报工作, 2014, 58(22): 38-47. (Yu Haiyan, Wei Junchao. Investigation and Analysis on University Data Curation Projects Abroad [J]. Library and Information Service, 2014, 58(22): 38-47.)
[36] Granger S. Emulation as a Digital Preservation Strategy [J]. D-Lib Magazine, 2000, 6(10). http://www.dlib.org/dlib/october00/granger/10granger.html.

作者贡献声明:

张梦霞: 政策细节梳理, 访谈专家学者, 论文起草;
顾立平: 研究设计, 信息源提供, 政策要素分析, 论文最终版本修订。

收稿日期: 2015-06-26
收修改稿日期: 2015-11-24

Policy Research of Data Curation

Zhang Mengxia[1,2] Ku Liping[1]
[1](National Science Library, Chinese Academy of Sciences, Beijing 100190, China)
[2](University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: [**Objective**] The implementation details of data curation promotes to establish data curation policy. [**Methods**] Based on reviewing the research achievements related to data curation, and then systematically concluded from the aspects as data selection standards, data storage standards, data communication and sharing mechanism. [**Results**] The main elements of data curation policy include: data selection standards (compliance with the requirements of data submission process, priority selection principle, the statement of data credibility and availability, data sources uncontroversial), data storage standard (follow the policy, guarantee data integrity, meet common technical standards, guarantee long-term sustainable development), and data communication and sharing mechanism (accordance with the laws and industrial directory, open access license, disclaimer propagation behavior, data reuse documentation). [**Limitations**] In the future, need to combine with the reality of China to complete the policy framework in detail. [**Conclusions**] Research organizations, associations and funding agencies should actively promote and develop data curation policy.

Keywords: Data curation Data management service Digital archives management Data rights management Data selection Long-term preservation